

Implementasi Metode *Fuzzy K-Means* untuk *Cluster* Judul Skripsi Mahasiswa

Ahmad Rifa'I¹, Galet Guntoro Setiadji²

¹ Sistem Informasi, Fakultas Teknolgi Informasi dan Komunikasi, Universitas Semarang,

² Teknik Informatika, Fakultas Teknolgi Informasi dan Komunikasi, Universitas Semarang,

¹ rifai@usm.ac.id, ² gallet@usm.ac.id

Abstract

Clusters are a way of classifying data which can later be used as information and processed using data mining methods with certain algorithms. From here the author wants to try to use data in the form of thesis titles or final assignments from students, which later can be grouped from these results. So that the existing data can be processed using a data mining method, namely *Fuzzy K-Means* (FKM).

The data used in this study uses thesis data of students majoring in information technology. As a comparison, the data is also processed using *K-Means*, from the *K-Means* calculation, the average cluster is 1.014 and the DBI validity is 0.725. Meanwhile, for the calculation of *Fuzzy K-Means*, the cluster average is 0.069 and the DBI validity is 0.304

Keywords: Data mining, K-Means, Fuzzy K-Means, DBI

1. Pendahuluan

Skripsi maupun tugas akhir menjadi salah satu syarat yang harus dipenuhi oleh mahasiswa untuk menyelesaikan studinya (Nurdin & Munthoha, 2017). Menyelesaikan skripsi maupun tugas akhir untuk mendapatkan gelar sarjana baik pada perguruan tinggi negeri maupun perguruan tinggi swasta merupakan suatu yang wajib dipenuhi oleh mahasiswa.

Era Industri 4.0 adalah nama tren otomasi dan pertukaran data terkini dalam hal teknologi industri kedepannya, sehingga adanya aplikasi sistem, teknologi *cloud*, dan internet semua serba berpacu tentang teknologi (Putrawangsa & Hasanah, 2018). Sehingga kita yang bekerja di dunia Pendidikan juga harus memulai mengikuti perkembangan teknologi.

Dimana sebagian mahasiswa khususnya fakultas teknologi informasi dan komunikasi mengarah ke sana. Mahasiswa biasanya akan membuat sebuah produk jadi berupa sistem aplikasi dan hardware. Dimana kita akan menyeleksi, yaitu dengan melihat judul skripsi atau tugas akhir mahasiswa.

Dari sana nanti kita akan kelompokkan dan diolah menggunakan sebuah metode data mining.

Dari hasil proses data mining dan menggunakan algoritma *Fuzzy K-Means* akan dibandingkan atau dikomparasi dengan metode *K-Means* nanti diharapkan akan bisa dikelompokkan bahwa

mahasiswa lebih sering membuat sebuah sistem atau hardware (Lestari, 2015).

Dan dari hasil komparasi dua algoritma yaitu *Fuzzy K-Means* dan *K-Means*, diharapkan akan mengetahui algoritma mana yang tepat dengan menggunakan data yang sudah disiapkan. Dengan bantuan software Rapidminer diharapkan dapat melihat nilai average dan nilai validitas menggunakan DBI dari masing-masing algoritma yang akan diuji nantinya.

Hasil dari penelitian ini diharapkan kurikulum akan lebih tepat sasaran dan menghasilkan lulusan mahasiswa yang siap bersaing di dunia kerja dengan Era Industri 4.0.

2. Metodologi

2.1. Data mining

Data mining merupakan proses mengali sebuah informasi dari kumpulan data-data, dari data itu nanti diolah menggunakan sebuah algoritma (Butarbutar, Windarto, & Hertama, 2016). *Data mining* banyak digunakan oleh peneliti atau sebuah perusahaan dalam menganalisa sebuah penjualan, yang akan menghasilkan sebuah strategi pasar penjualan. Sehingga sebuah perusahaan bisa memperkirakan atau memprediksi sebuah laba atau rugi.

Data mining memiliki tahapan proses atau pola mengolah data, yaitu :

- Pembersihan data (menghapus data-data yang tidak digunakan, dan mengisi data yang hilang)
- Integrasi data (menjadikan satu data-data dari berbagai sumber)
- Pemilihan data (memilih data-data yang sesuai dan relevan)
- Transformasi data (mengubah data kedalam format tertentu agar bisa diolah menggunakan algoritma datamining)
- Evaluasi pola (mengenali pola datamining yang sesuai dengan data)
- Penyajian pola (memvisualkan sebuah hasil olah datamining)

2.2. K-Means

K-Means adalah adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya (Rahmawati, Widya, & Suryani, 2014). Alur perhitungan menggunakan metode *K-Means* adalah sebagai berikut :

- Pilih K buah titik centroid secara acak
- Kelompokkan data sehingga terbentuk K buah cluster dengan titik centroid dari setiap cluster merupakan titik centroid yang telah dipilih sebelumnya.

$$d_{ij} = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]}$$

Perbaharui nilai titik centroid

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

- Ulangi langkah 2 dan 3 sampai nilai dari titik centroid tidak lagi berubah

2.3. Fuzzy K-Means Clustering (FKM),

Fuzzy K-Means Clustering (FKM), atau dikenal juga sebagai *Fuzzy Isodata* merupakan salah satu metode clustering yang merupakan bagian dari metode *Hard K-Means*. *Fuzzy K-Means* menggunakan model pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0

hingga 1. Tingkat keberadaan data dalam suatu kelas atau *cluster* ditentukan oleh derajat keanggotaannya (Mas'udin, Farida, & Mustafa, 2018).

Konsep dasar *Fuzzy K-Means*, pertama kali adalah menentukan pusat *cluster* yang akan menandai lokasi rata-rata untuk tiap-tiap cluster. Pada kondisi awal pusat *cluster* ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap cluster. Dengan cara memperbaiki pusat cluster dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat cluster akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat cluster yang berbobot oleh derajat keanggotaan titik data tersebut.

Untuk tahapan FKM hampir sama dengan *K-Means*, dengan tahapan sebagai berikut:

- Menentukan awal cluster K
- Menentukan nilai koefisien secara acak di setiap titik data cluster

$$\sum_{j=1}^k u_{ij} = 1$$

- Ulangi algoritma perhitungan 4 dan 5 hingga konvergen atau tidak ada perubahan pada nilai iterasi dan nilai ambang batas centroid
- Melakukan perhitungan centroid dari masing-masing kluster

$$0 < \sum_{i=1}^n u_{ij} < n$$

- Hitung koefisiennya di tiap cluster.

$$u_{ij} = \frac{D(x_i, c_j)^{-\frac{2}{w-1}}}{\sum_{l=1}^k D(x_i, c_l)^{-\frac{2}{w-1}}}$$

2.4. *Davies Bouldin Index* (DBI) diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 adalah metrik untuk mengevaluasi hasil algoritma clustering. Adapun Langkah-langkah dalam proses perhitungan DBI adalah sebagai berikut :

- SSW (sum of within cluster)

$$SSW_i = \frac{1}{m_1} \sum_{j_1}^{m_1} d(x_j, c_i)$$

- SSB (sum of square between)

$$SSB_{i,j} = d(c_i, c_j)$$

c. Ukuran rasio (R) dan DBI

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

3. Hasil dan Analisa

Data yang akan di gunakan dalam penelitian merupakan data judul skripsi yang ada di Program Studi Teknik Informatika dan Sistem Informasi, data di ambil mulai dari Tahun lulus 2011 sampai dengan 2019. Kemudian dilakukan normalisasi data, dan menghasilkan data seperti ditabel Tabel 3.1. Tabel Judul Skripsi berikut ini.

Tabel 3.2. Tabel Judul Skripsi

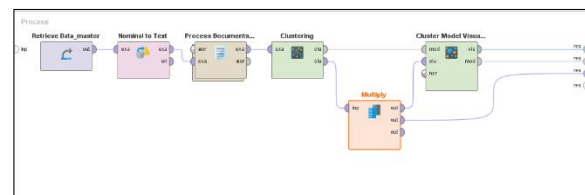
No	Judul Skripsi
1	Memahami Komunikasi Keluarga dalam Proses Penguatan Hubungan Antar Anggota Keluarga Besar
2	Sistem Pendukung Keputusan Pemilihan Karyawan Teladan Pada PT. Jawa Surya Kencana Indah Menggunakan Metode Simple Addictive Weighting
3	Analisis Wacana Berita Dugaan Kasus Chat Pornografi Yang Dilakukan Oleh Habib Rizieq Pada Metro TV
4	Analisis Wacana Berita Kasus Dugaan Penistaan Agama Ahok di Viva.Co.Id
5	Pengaruh Program Tayangan Kontes Dangdut D'Academy di TV Swasta Indosiar dengan Minat Siswa Non Dangdut Sekolah Musik Adsor Purnomo Semarang pada Musik Dangdut
6	Proses Komunikasi Organisasi Komunitas Boomerang Sport Semarang Dalam Memperkenalkan Olahraga Ekstrem Boomerang yang Sarat Akan Edukasi dan Prestasi di Semarang
....
3447	Strategi Komunikasi Eksternal Public Relation Dalam Membentuk Citra Program USM Fiesta

Didalam penelitian ini kita akan mencoba menggunakan Fuzzy K-Means (FKM), namun FKM untuk perhitungan algoritmanya hampir sama dengan K-Means, namun di tahapannya FKM menggunakan perhitungan Fuzzy. Sebelum nya penelitian ini menggunakan software Rapid minner, sebelum menghitung Fuzzy K-Means kita akan melakukan perhitungan menggunakan metode K-Means yang nanti akan kita bandingkan dengan FKM.

3.1. Tahapan K-Means

Berikut ini tahapan-tahapan dan gambaran penerapan K-Means menggunakan rapidminer. Dimana operator data retrieve data_master sebagai data, dan operator nominal to text yang mengubah nilai nominal text.

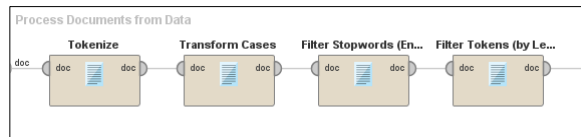
Kemudian baru process document dimana memfilter kalimat menjadi beberapa kata yang kemudian akan di proses di operator cluster, dan multiply dan dapat dilihat pada Gambar 3.1. Tahapan Penerapan K-Means dibawah ini.



Gambar 3.1. Tahapan Penerapan K-Means

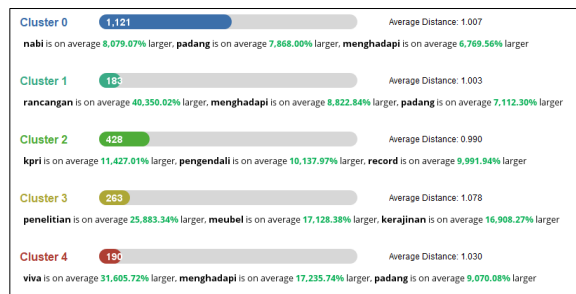
Tahap selanjutnya yaitu operator model yang akan menghasilkan validitas DBI dan Head Map.

Dan selanjutnya berikut ini step-step operator yang ada didalam processing document dari data text seperti pada Gambar 3.2. Proses Dokumen dari Data Teks dibawah ini.



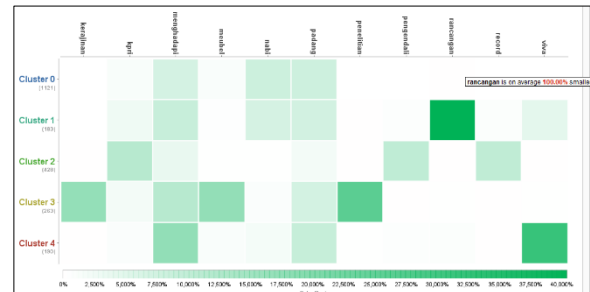
Gambar 3.1. Proses Dokumen dari Data Teks

Dari penggunaan metode k-means di rapid minner dengan 5 clusster didapatkan data sebagai berikut : menggunakan perhitungan jarak terdekat menggunakan Euclidean Distance didapatkan average cluster distance sebesar 1.014 dengan validitas DBI sebesar 7.255. Sedangkan hasil akhir dari 5 cluster setelah dihitung dengan rapidminer seperti gambar dibawah ini.



Gambar 3.2. Hasil Akhir

Dari hasil head map rapidminner didapatkan data sebagai berikut, dengan head map di cluster 1 averagennya sebesar 40.350,02% dengan text yang mengandung kata “rancangan”.



Gambar 3.3. Head Map Rapidminner

Sehingga apabila kita mengecek sesuai dengan judul yang kita gunakan akan terdapat 9 judul yang mengandung kata “rancangan”, seperti pada tabel 3.2. Head Map Rapidminner.

Tabel 3.2. Head Map Rapidminner

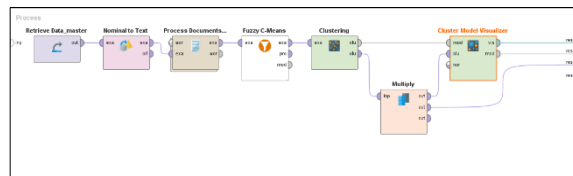
NIM	NAMA	JUDUL
G.131.13.0050	Mallinda Nurul Ichani	Rancangan Bangun Sistem Informasi Pembayaran SPP pada SD Islam Terpadu Bunayya Semarang Berbasis Web
G.231.08.0039	Pratiknyo	Perancangan Laboratorium E-labs PT. Rasapala Semarang
G.231.11.0031	Bungas Septian Sarwo Edhi	Analisis Dan Perancangan Management Network PT. Angkasa Pura I Semarang Menggunakan Mikrotik
G.211.13.0138	Zulfan Effendy Perdana	Analisis Pengaruh UI/UX terhadap Desain Interaksi Perancangan dan Pengembangan Website Sistem Informasi Akademik Universitas Semarang (SIA USM)
G.211.13.0089	Agung Prabowo	Perancangan Dan Pembangunan Sistem Manajemen File Pada Fakultas Teknologi Informasi Dan Komunikasi Universitas Semarang Menggunakan Framework PHP Laravel Dan Database Maria DB

NIM	NAMA	JUDUL
G.231.12.0021	Kiki Baehaki	Perancangan dan Pembuatan Load Balancing Pada Web Server Sistem Informasi Bimbingan Mahasiswa Menggunakan Apache
G.211.14.0159	Rizki Setiawan	Perancangan Sistem Informasi Pengolahan Nilai Siswa Secara Online Pada SD Negeri Susukan 04
G.211.11.0005	Remyta Faried Yunus	Perancangan Sistem Perwalian Berbasis Web
G.231.12.0033	Djayus Wijaya	Rancangan Anggaran Biaya Menggunakan Visual Basic 2008 Pada CV Cemerlang Semarang

Kalau kita lihat dengan penggunaan algoritma K-Means hampir sebagian besar bahwa mahasiswa lebih cenderung mengarah mengambil aplikasi online dengan menggunakan web sebagai salah satu judul skripsinya

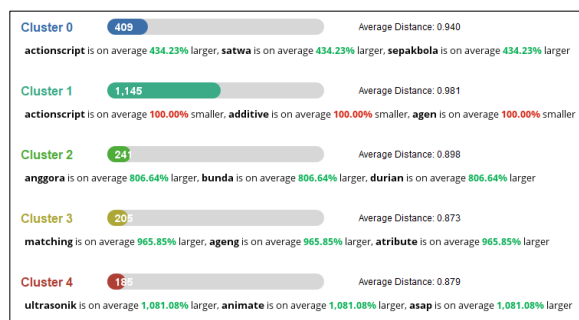
3.2. Penerapan Fuzzy

Berikut ini screen shoot penerapan Fuzzy K-Means dengan menggunakan tools Rapid Miner, dimana tahapannya hampir sama dengan K-Means namun ada penambahan model fuzzy didalam tahapannya seperti gambar dibawah ini.



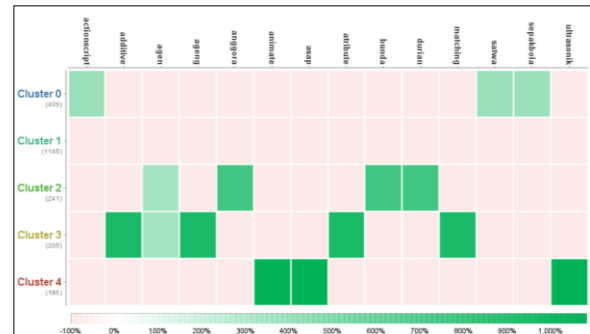
Gambar 3.4. Tambahan Model Fuzzy

Proses Fuzzy K-Means hampir sama dengan K-Means, dan dari 5 cluster menghasilkan data sebagai berikut : penggunaan perhitungan jarak terdekat menggunakan Euclidean distance sebesar 0.945 dan validitas DBI sebesar 0.695. Dari hasil 5 cluster didapatkan cluster terakhir seperti gambar dibawah ini



Gambar 3.5. Hasil 5 Cluster

Dari hasil head map rapidminier menggunakan FKM didapatkan, bahwa cluster 5 memiliki head map average sekitar 1.081,08% dengan text yang mengandung kata “animate, asap dan ultrasonik”.



Gambar Error! No text of specified style in document..6. Head Map pada Fuzzy K-Means

Judul-judul yang mengandung kata animate, asap dan ultrasonik dapat dilihat pada tabel 3.3. Daftar Kata animate, asap, ultrasonic dapat dilihat pada tabel dibawah.

Dalam hasil perhitungan Fuzzy K-Means menggunakan Rapidminer dan apabila kita lihat mahasiswa lebih cenderung mengambil tema berbasis hardware dan software seperti penggunaan Arduino dan android dalam menyelesaikan judul skripsinya.

Tabel 3.4. Tabel 3.3. Daftar Kata animate, asap, ultrasonic

1	Kata untuk Animate		
	Nim	Nama	Judul
	G.231.12.0054	REZA ARI PERDANA	Rancang Bangun Aplikasi Panduan Menghadapi Bencana Alam Berbasis Android Menggunakan Adobe Animate CC
	G.211.12.0178	MUH AFIF NOORHIDAYAT	Rancang Bangun Aplikasi Pembelajaran Hardware Komputer Berbasis Android Menggunakan Adobe Animate CC
2	Kata untuk Asap		
	Nim	Nama	Judul
	G.231.08.0039	PRATIKNYO	Perancangan Laboratorium E-labs PT. Rasapala Semarang
	G.211.14.0018	Ryan Sofyan	Rancang Bangun Alat Filter Asap Dapur Otomatis Deteksi Kebakaran Dengan Sensor MQ-2 dan Flame Detector Menggunakan Arduino
	G.231.11.0092	CHOIRUL ANWAR	Rancang Bangun Detektor Temperatur Suhu di Kamar Asap Kebun Ngobo PT. Perkebunan Nusantara IX Berbasis Iot
3	Kata untuk Ultrasonik		
	Nim	Nama	Judul
	G.211.14.0125	RENDY HARTONO	Rancang Bangun Alat Bantu Parkir Kapal Menggunakan Sensor Ultrasonik Berbasis Arduino
	G.211.14.0150	ADINA NUR FARIDA	Rancang Bangun Robot Pencari Jalan Keluar Berbasis Arduino Dengan Sensor Ultrasonik Menggunakan Metode Fuzzy Logic
	G.211.15.0090	SANDRA	Simulasi Rancang Bangun Alat Pendeteksi Ketinggian Air Rob di Citarum Menggunakan Ultrasonik Berbasis IOT Dengan Nodemcu ESP8266

Dalam hasil perhitungan Fuzzy K-Means menggunakan Rapidminer dan apabila kita lihat mahasiswa lebih cenderung mengambil tema berbasis hardware dan software seperti penggunaan Arduino dan android dalam menyelesaikan judul skripsinya.

Daftar Pustaka

- Butarbutar, N., Windarto, A. P., & Hertama, D. (2016). Komparasi Kinerja Algoritma Fuzzy C-Means dan K-Means dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai akademik Siswa. *Jurasik*, 46-55.
- Lestari, W. (2015). Pemetaan Gaya Belajar Mahasiswa dengan Clustering Menggunakan Fuzzy C-Means. *Jurnal Sainstech Politeknik Indonusa Surakarta*, 1-8.
- Mas'udin, P. E., Farida, A., & Mustafa, L. D. (2018). ClusteringData RemunerasiDosen Untuk Penilaian Kinerja Menggunakan Fuzzy c-Means. *Jurnal Resti*, 288-294.
- Nurdin, N., & Munthoha, A. (2017). Sistem Pendeteksi Kemiripan Judul Skripsi Menggunakan Algoritma Winnowing. *Jurnal Nasional Informatika dan Teknologi Jaringan*, 90-97.
- Priliant, K. R., & Wijaya, H. (2014). Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering. *Jurnal Cybermatika*, 1-6.
- Putrawangsa, S., & Hasanah, U. (2018). Integrasi Teknologi Digital dalam Pembelajaran di Era Industri 4.0. *Jurnal Tatsqif*, 42-54.
- Rahmawati, L., Widya, S. S., & Suryani, E. (2014). Analisa Clustering Menggunakan Metode K-Means dan Hierarchical Clustering (Studi Kasus : Program Studi Kimia). *Jurnal Teknologi dan Informasi*, 1 - 8.
- Riyadhi, M. F. (2019). Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi Dengan Metode K-Means Clustering (Studi Kasus: Prodi Sistem Komputer). *Jurnal Komputika*.
- Windarto, A. P., & Hertama, D. (2016). Komparasi Kinerja Algoritma Fuzzy C-Means Dan K-Means dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai akademik Siswa. *Jurasik*, 46-55.
- Lestari, W. (2015). Pemetaan Gaya Belajar Mahasiswa dengan Clustering menggunakan Fuzzy C-Means. *Jurnal Sainstech Politeknik Indonusa Surakarta*, 1-8.
- Mas'udin, P. E., Farida, A., & Mustafa, L. D. (2018). ClusteringData RemunerasiDosen Untuk Penilaian Kinerja Menggunakan Fuzzy c-Means. *Jurnal Resti*, 288-294.
- Putrawangsa, S., & Hasanah, U. (2018). Integrasi Teknologi Digital dalam Pembelajaran di Era Industri 4.0. *Jurnal Tatsqif*, 42-54.